Journal of Archaeological Science 30 (2003) 1685–1693

Journal of Archaeological SCIENCE

http://www.elsevier.com/locate/jas

Palaeolithic radiocarbon chronology: quantifying our confidence beyond two half-lives

P.B. Pettitt [a]*, W. Davies [b], C.S. Gamble [b], M.B. Richards [c]

[a] Department of Archaeology, University of Sheffield, Northgate House, West Street, Sheffield S1 4ET, UK
[b] Department of Archaeology, University of Southampton, Avenue Campus, Southampton SO17 1BJ, UK
[c] Department of Chemical and Biological Sciences, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK

Received 18 September 2002; received in revised form 6 May 2003; accepted 6 May 2003

Abstract

It is now three decades since Waterbolk introduced evaluation criteria to ^{14}C chronology. Despite this, and other subsequent attempts to introduce quality control in the use of ^{14}C data, no systematic procedure has been adopted by the archaeological community. As a result, our databases may be significantly weakened by questionable dates and/or questionable associations between dated samples and the archaeological phenomena they are intended to represent. As the use of chronometric data in general becomes more ambitious, we must pause and assess how reliable these data are. Here, we forward a set of evaluation criteria which take into account archaeological (e.g. associational, stratigraphic) and chronometric (e.g. pre-treatment and measurement) criteria. We intend to use such criteria to evaluate a large ^{14}C dataset we have assembled to investigate Late Glacial settlement in Europe, the Near East and North Africa, supported by the Leverhulme Trust. We suggest that the procedure presented here may at least form the basis of the development of more rigorous, scientific use of ^{14}C dates.

Keywords: ^{14}C; Radiocarbon; Chronometry; Late Glacial; Palaeolithic

1. Introduction

^{14}C dating has played a central role in the archaeological research landscape for over fifty years, and during that time many thousands of ^{14}C measurements have been obtained for archaeological sites. During the last three decades these have been increasingly supplemented by results from other radiometric techniques. While it is generally accepted that some ^{14}C dates have greater archaeological validity than others, archaeologists often seem reluctant to be fully explicit about their selection criteria for retaining or rejecting determinations. This has resulted in an abundance of questionable ^{14}C measurements and conclusions, which appear time and time again in the literature. As the questions we are asking of the data become more ambitious, addressing for example human demographics and responses to rapid environmental changes, there is clearly a need to evaluate and refine our chronological data. The rigorous attempts by Spriggs [28] and Spriggs and Anderson [29] to inject ‘chronological hygiene’ into the dating of the Southeast Asian island Neolithic and colonisation of East Polynesia respectively serve as admirable examples of what archaeologists should be doing with their dates as a matter of routine. This paper therefore, does not pretend to be the first to propose explicit selection criteria for ^{14}C determinations. Rather, it is an initial probe into a more rigorous treatment of dates and we hope that it will stimulate a wider debate amongst archaeologists about how we set about rationalising our burgeoning databases. From the outset we do not advocate the suppression of dates deemed to be “inadequate” for archaeological analyses, but do ask that our working databases be reduced to a more manageable “rump” of verified determinations. Here, we propose selection criteria specifically designed for application to a dated-site

* Corresponding author. Tel.: +44-1865-273931; fax: +44-1865-273932.
E-mail address: paul.pettitt@keb.ox.ac.uk (P.B. Pettitt).

0305-4403/03/$ - see front matter
doi:10.1016/S0305-4403(03)00070-0

database under construction to examine human movements in Europe, the Near East and North Africa between 20 and 8 ka BP (i.e. between ~1 and ~4 half-lives of ^{14}C), funded by the Leverhulme Trust. The criteria we employ fall into two broad categories—chronometry (i.e. methodological issues) and interpretation (i.e. archaeological issues). While the former is the domain of radiocarbon specialists and the latter of archaeologists, a meaningful discourse between the two is crucial. We intend to grade each age determination in the database according to the criteria published below, with refinements from further debate that we hope this paper will stimulate. Although the issues described and discussed below are designed to work for our own database, we hope that they will have a wider significance for researchers in other archaeological periods as well, and welcome responses towards refining the evaluation criteria.

2. Using ^{14}C dates: problems historical and current

Three decades ago, Waterbolk advanced a set of propositions by which the reliability of radiocarbon dates could be ascertained, with a view to "…improving the utilisation of ^{14}C dates in archaeology" [35, p. 15]. The increasing desire and ability to calibrate makes these propositions and the problems they address even more pertinent today. He addressed nine potentially problematic areas:

1. The certainty of association between a dated sample and the archaeology/event that it is intended to date.
2. The difference in age between the sample and the date of its deposition, e.g. 'old wood' effects.
3. The contamination of samples with younger or older carbon bearing materials such as humic acids and carbonates.
4. The differential effects of contamination depending on sample age (i.e. the older a sample the greater the potential effect).
5. Potential problems with certain chemical fractions, e.g. the relatively open system of bone being more susceptible to contamination and the question of burnt bone.
6. Inter laboratory pretreatment and measurement error.
7. The question of 'averaging' dates from large data sets.
8. The interpretation of large data sets.
9. The issue of calibration, which at the time (1971) was applicable back to c. 5000 BC.

Two years after Waterbolk's clarification of archaeological and methodological issues, Renfrew published his account of the two 'revolutions' in ^{14}C dating, i.e. the discovery of the potential (in 1947) and practicability (in 1948) of the technique itself and the impact of calibration. He noted the four main assumptions underpinning the technique, namely (a) the half-life of ^{14}C now known to be ~5730 years, (b) the absence of contamination, (c) the uniform worldwide distribution of ^{14}C and (d) a consistent production of ^{14}C in the upper atmosphere over time. At the time of writing it was clear that contamination, then of the large and often bulked samples required for conventional measurement methods, was an ever-present danger. A small degree of latitudinal differences in ^{14}C mixing had been observed, and that major variations in ^{14}C pathways between terrestrial and marine biotopes was understood to have potentially major effects on ^{14}C dates (e.g. [31, p. 1045]). It was also clear that Libby's assumption that atmospheric production levels had remained stable was incorrect.

Despite a third 'revolution' in the technique, i.e. that of Accelerator Mass Spectrometry by which the majority of ^{14}C measurements are made today, many of the problems raised by Waterbolk and Renfrew remain current. Here, we consider these issues, and develop from these evaluation criteria as discussed above. We consider the auditing of absolute determinations under two main headings: (a) methodological concerns (Waterbolk's issues II–IX) and (b) archaeological ones (Waterbolk's issue I, with additions by ourselves). Our proposed grading system aims to treat both these concerns separately and then in tandem, with the concerns of methodology probably being found less contentious than ones of archaeological significance and meaning.

It seems to be implicitly acknowledged by researchers (e.g. [1,14,35]) that it is easier to quantify levels of confidence in absolute dates from a methodological perspective than an archaeological one. In practice, it is easier to identify and deal with chronometric errors, as indicators of contamination exist such as the stable isotopes of carbon and nitrogen, as do indicators of problematic or erratic measurement itself. If any of these arise, the sample or resulting 'date' can simply be eliminated. By definition, the publication of a ^{14}C date accompanied by a laboratory number should be a clear statement from the laboratory that, *as far as was ascertainable from the data at hand* the pretreatment and measurement were unproblematic between accepted assessment parameters. The relatively simple and logical means by which chronometry can be assessed is reflected in the number of criteria (II–IX) that Waterbolk identifies for assessing the methodological reliability of absolute determinations. By contrast, the difficulty of approaching issues of interpretation in a similar, logical manner is reflected by his solitary archaeological criterion (I): a ratio of eight to one.

Somehow, archaeologists must grasp the nettle and place less reliance on intuitive responses to the validity of absolute dates and build into their use of them an

ongoing critique. We acknowledge that this task is difficult, but it should not be abandoned for this reason, and is not impossible if we are open about the decisions we take and how we make them explicit. It is important that we do not unwittingly place greater emphasis upon our relative archaeological chronologies (derived from typological and technological variability) than on absolute dated ones: we may be discarding methodologically valid dates simply because they disagree with our preconceptions about the development of cultural sequences (e.g. [8,36]). Stuart Piggott, for example ([23]: see also [24]), famously dismissed surprisingly early dates for British Neolithic monuments as "archaeologically inacceptable" as they did not agree with his setting of the beginning of the Neolithic at 2000 BC. Such attitudes surely defeat the primary purpose of having absolute dates, which is to provide an independent chronological check on our relative archaeological chronologies. If one takes the latter view, one is left with two possible views on "aberrant" absolute dates: either they have been moved stratigraphically, or our current views of archaeological developmental change are too rigid and take little account of true assemblage variability in the Palaeolithic record.

Questions about the archaeological significance of absolute dates are of course difficult to divorce from methodological ones, although the reverse is less certain. The major archaeological concern must surely be the stratigraphic context of a dated sample. Conventional ^{14}C methods often relied on bulked samples, which ran the risk of effectively averaging materials (anthropologically modified or otherwise) from more than one carbon source. With the rise of AMS ^{14}C dating in the 1980s the question of association between dated samples and behaviourally diagnostic archaeology was reduced, as the small sample sizes required made the direct dating of relevant materials possible, although different archaeological concerns then attained significance, notably the stratigraphic mobility of small samples.

It would be comforting to think that the question of association between samples selected for ^{14}C dating and their related objects had long been put to rest. Many scholars now rely only on archaeological materials bearing true signs of hominin manufacture or modification as reliable (e.g. [2,15,30]). This is sadly not always the case, however, and a number of important behavioural issues are debated using dates on charcoal fragments for which we cannot eliminate natural causes, however unlikely. Other more questionable areas remain, and it is sad that over thirty years later some of Waterbolk's concerns are as pertinent as ever. We regard as still potentially problematic Waterbolk's issues III (contamination), IV (differential effects of contamination by age, an issue we regard as part of the wider concern of coarse and heterogeneous precision over two half lives), V (chemical fractions), VI (inter laboratory errors), VIII (interpretation of large data sets) and IX (calibration, which we consider to be part of the wider issue of accuracy).

Here, we have taken the issues raised by Waterbolk which we believe still have currency, and have modified and added to them. We evaluate each sample in terms of our criteria on a point basis, beginning from 0 (reflecting very poor confidence in the aspect of concern) to 4 (very high confidence). The resulting 'scores' provide a reflection of the reliability of the date and its relevance to archaeological issues. We realise the arbitrary nature of such a procedure. Given this, it seemed logical to arbitrarily chose 40% (i.e. scores of 0 and 1) as a cut off point below which were have little or no confidence in the attribute of concern, 40–60% (score of 2) as falling into a category of questionable confidence and 60% or above (scores of 3 or 4) as reflecting confidence. We combine individual scores into an overall evaluation score which uses the same cut-off points.

3. Evaluation criteria

3.1. Chronometry

3.1.1. Contamination by older/younger carbon and measurement of irrelevant carbon fractions

A sample is contaminated if its ^{14}C/^{12}C ratio has changed since deposition by any process other than radioactive decay [10]. With the small amounts of residual ^{14}C in samples beyond 2 or 3 half-lives, a very small amount of residual contamination from an irrelevant carbon source may have drastic effects on the resulting measurement. In addition to this, measuring very small samples of carbon, i.e. from samples where taking larger samples is impossible or from those in which carbon preservation has simply been low, raises the issue of residual contamination further, and will produce relatively large laboratory errors, as may render the measurements of limited use. While indicators of potential contamination may occur during pretreatment or measurement, such as an erroneously high nitrogen content (and therefore C/N ratio) contamination may still occur without obvious chemical indicators. In addition to this, samples that may contain carbon from numerous sources, e.g. rock art pigments, while not strictly speaking contamination, do raise the issue of the relevance of the dated fraction to the archaeological issue at hand. Our evaluation therefore builds in uncertainty in this area.

1. Carbon derives from a questionable chemical fraction, e.g. burnt bone, humic acid, oxalate crust, apatite, or the C/N ratio indicates potential contamination.
2. Amount of carbon measured too small to allow C/N evaluation.
3. Carbon derives from a chemically complicated sample material from which numerous carbon

sources cannot be ruled out by standard pretreatment methods, e.g. rock art pigments, but that otherwise appears unproblematic, or from a sample with an unknown conservation history.
4. Carbon derives from collagen from bone, antler or ivory for which pretreatment data are unproblematic.
5. Carbon derives from specific amino acid known to grow only in bone, or from the wood charcoal fraction of a charcoal sample identified to genus and for which the 'old age' effect can be eliminated.

3.1.2. ^{14}C dating of different chemical fractions

Mellars [18,19] has suggested that due to methodological issues measurements obtained from charcoal samples may yield relatively older dates than those on bone samples which are chemically more open systems and therefore potentially open to contamination. There is no a priori reason why this should be so, and dating of clearly associated bone/charcoal pairs demonstrates that this is probably a minor issue, but a potential problem must be admitted. Pettitt [21], for these reasons, has suggested caution over the interpretation of dates from the Pavlovian of the Czech republic (mainly on charcoal) which appear to be generally earlier than those of the mid Upper Palaeolithic Russian Plain (mainly on bone) as suggesting a population movement in response to environmental deterioration.

1. Measurements were on samples of the same material, and fall outside of a chronological sequence taking into account layers above and/or below.
2. Measurements were on samples of the same material but no available crosscheck with other dated horizons is available.
3. Measurements were on samples of the same material but these are in agreement with samples of other materials from above or below the relevant horizon, i.e. fall into a clear sequence.
4. Measurements were upon at least a pair of charcoal (1 measurement) and one bone/antler/ivory pair, which were statistically the same age at 2σ[1].
5. Measurements were upon several discrete materials, at least one of which was charcoal and one of which was bone/antler/ivory, all clearly in association and statistically the same age at 2σ.

3.1.3. Accuracy

The accurate measurement of ^{14}C, present at 10^{-12} to 10^{-15} of the level of ^{12}C, has always been problematic (e.g. [12,13]). Fluctuations in the atmospheric composition of carbon, and deviation of certain samples from equilibrium with the atmosphere, severely affect accuracy and in some cases are still poorly understood. From early in the technique's history the fluctuation of ^{14}C production in the upper atmosphere as revealed by the ^{14}C and dendrochronological dating of tree rings, was seen as problematic. The apparent magnitude of such fluctuations for beyond ~3 half-lives is now becoming increasingly apparent in comparisons of ^{14}C and Uranium-Series dating of flowstone and coral samples (e.g. [17,25,32]). While calibration curves such as the internationally accepted INTCAL98 [31] now take us to ~25,000 BP, the number of data points before ~12,550 BP (~15,000 cal BP) are small and the data highly problematic beyond ~30,000 BP. In addition to atmospheric effects, the mixing of carbon from different reservoirs—notably the deep-ocean and surface waters generally referred to as the 'marine reservoir effect'—will cause animal samples from such environments to deviate from atmospheric equilibrium. Correction for such effects is possible, although for some reservoirs such as rivers the amount of required correction is still unknown. Where correction curves have been constructed, different isotopic fractionation region to region may call into question the appropriateness of certain curves. Accuracy is still very much a major issue.

1. Sample dates to >30,000 BP and is the solitary date for a given horizon or falls outside of a sequence of dates from horizons above and/or below that from which it came.
2. >2 samples from a given horizon date to >30,000 BP and are generally in agreement with a chronological sequence with no more than 1/6 dates as outliers.
3. >2 samples from a given horizon date to >30,000 BP but fall into a clear chronological sequence with few or no outliers.
4. Samples date to <30,000 BP, fall into a clear chronological sequence with few or no outliers, and/or may be calibrated using INTCAL98.
5. Samples date to <20,000 BP, fall into a clear chronological sequence and/or may be calibrated using INTCAL98.

3.1.4. Sample materials and ^{14}C measurement

Most dated carbon derives from general collagen in organic samples or from wood carbon in charcoal. In some cases however, other fractions have been dated, either through choice or through the unavailability of more suitable fractions. These are often problematic. Similarly, marine reservoir effects or old wood samples can cause over-estimation of ^{14}C ages and therefore feed directly into issues of accuracy as noted above. The agreement of measurements on different chemical fractions of the same sample—e.g. humic and wood charcoal (humin) fractions of charcoal samples—will allow far greater confidence in the result than measurements that differ.

[1] We suggest that all discussions of date ranges use 2σ only, given the greater probability (i.e. 95%) that the true age of the sample lies within this range.

1. Sample is of riverine or marine derivation, and has not been corrected for a reservoir effect *or* sample is of wood charcoal and clearly not of a twig or small branch, which has not been identified to genus and for which therefore an 'old age' overestimation cannot be ruled out. Also, two chemical fractions of the same sample have been measured and differ at 2σ.
2. Carbon measured derived from a problematic chemical fraction, e.g. apatite, humic acids, carbonates, *or* carbon measured is exceptionally low, i.e. <0.5 mg *or* C/N ratio is outside of acceptable range for sample material (e.g. see [6] for consensus values).
3. Collagen/cellulose yield and/or carbon yield relatively low (e.g. >0.5 mg carbon measured), but otherwise unproblematic.
4. Respectable yields from collagen/cellulose fractions and no indicators of pretreatment problems.
5. As 3, and sample is statistically the same age as samples of *other materials* from same stratigraphic horizon.

3.1.5. Sample measurement and reporting

Pretreatment and measurement techniques have undergone considerable refinement over the last 10–20 years, in addition to the refinement that the AMS technique brought in itself. Respectable laboratories ensure consistency and standards by participating in informal cross testing and the formal International Radiocarbon Intercomparison. Whilst results measured during the early days of the technique or by laboratories that do not participate in such intercomparisons are not necessarily wrong, uncertainty in this area should be taken into account. In addition, the credibility of measurements may be brought into question if they are not supported in print by laboratory analytical data such as C/N ratios (in the case of bone collagen). We have little confidence in measurements of bulked samples for which one cannot eliminate the contribution of several carbon sources. While measurements on such samples may be a true reflection of carbon isotopes present, the resulting age may be a meaningless combination of these multiple carbon sources. We automatically score bulked samples at 0. By contrast, 'single entity' samples, from which the dated carbon derives from one reservoir assuming pretreatment has successfully removed contamination, are far more reliable than bulked.

1. Sample was created from a bulked sample and/or measured conventionally before 1970.
2. Sample was pretreated and/or measured at a laboratory that does not participate in International Radiocarbon Laboratory intercomparisons.
3. Sample measurement is published without pretreatment and measurement methods, or no laboratory comment that results satisfied the laboratory's assessment criteria.
4. Sample is published with such data, although some criteria fall outside of acceptable limits.
5. Sample is published with full pretreatment, measurement and stable isotope data, all of which satisfy accepted criteria.

3.2. Interpretation

3.2.1. Certainty of association of dated sample with human activity

Unless dated samples are of unquestionable human manufacture, there will always be an issue over the degree of confidence that they reflect human, as opposed to animal or other non-human activity.

1. Low possibility (very poor archaeology, item recovered from mainly palaeontological (e.g. denning) horizon.
2. Reasonable possibility (archaeology scattered and/or fragmentary, low numbers).
3. Probability (no demonstrable relationship but number of items and spatial patterning suggest association).
4. High probability (direct functional/contextual relationship).
5. Full certainty (anthropogenic object of concern dated).

3.2.2. Relevance of dated sample to specific archaeological entity of concern

Given the geologically complex nature of Pleistocene (and many Holocene) sites, notably caves and rock shelters, simply demonstrating confidently that the dated samples reflect human activity need not demonstrate that this activity pertains to the specific *cultural* remains of concern. For example, a bone fragment bearing a stone tool cut mark recovered from an Aurignacian horizon, whilst clearly dating human activity, need not reflect *Aurignacian* activity, but could relate to depositional, post-depositional mixing or interstratified visits by different cultural or biological groups. This issue is especially relevant to possible Neanderthal and modern human overlap, e.g. at the Grotte du Renne, Arcy-sur-Cure ([16]; cf. [20]) and level G_1 at Vindija Cave, Croatia (e.g. [27]).

1. Sample material (or genus if charcoal) is unknown.
2. No existing/published traces of hominin manufacture or modification of sample object exist.
3. Sample has high association with diagnostic archaeology, through incorporation in same horizon/level, but is in itself undiagnostic.
4. High probability of association, through incorporation into clear feature, e.g. hearth, pit, channel, very discrete occupation horizon, albeit undiagnostic itself.

5. Sample dated is either culturally diagnostic itself (or a hominin fossil), or bears *both* a high probability of association in addition to clear traces of hominin manufacture/modification.

3.2.3. Quantity and nature of dates for archaeological horizon

One date is no date, given that it is impossible to evaluate whether or not it is correct. Interpretation is similarly difficult if the only dates that exist are statistically distinct at 2σ. Palimpsest sites are usually complex stratigraphically and one may assume that processes of deflation have often confused the chronological issue. Given this, it is not surprising that a certain number of dates in a given sequence or stratum will be outliers for no methodologically or archaeologically obvious reasons. By 'outlier' we refer to dates that are statistically distinct from the main group/sequence at 2σ. It follows that the larger amount of statistically identical dates available for a given horizon the more confidence one may have in the resulting ages. Colleagues that have identified gross outliers may want to eliminate them automatically, although we recognise that this procedure may remove good dates. This having been said, we appreciate that uncalibrated outliers may disappear upon calibration, and that outliers should be investigated by statisticians (Buck, pers. comm.), but given the infancy of calibration/correction for radiocarbon dates beyond two half-lives we err on the side of caution here. Datasets that are calibrated in future can, after all, be re-integrated into analyses. We appreciate that if corroboration of dates by agreement with others is not possible, the date will get a relatively low score. While such dates may not be problematic, and while their credibility in this criterion is not on a par with a date, for example, on burnt bone, there is a progression of confidence with corroborative dates and we err on the side of caution.

1. The date is the sole measurement for a given horizon, or is one of several that differ statistically at 2σ.
2. The date is one of only 2 dates for a given horizon which are statistically the same age at 2σ.
3. The date is one of a group of >2 dates for a given horizon which are statistically the same age at 2σ.
4. The date is one of >3 dates for a given horizon, which are statistically the same age at 2σ.
5. The date is one of >5 dates for a given horizon which are statistically the same age at 2σ.

3.2.4. Sample materials and stratigraphic issues

It soon became obvious in the infancy of AMS radiocarbon dating that the relatively small samples available for dating could well be stratigraphically mobile. Sample selection should, but often does not, control for this possibility. Ideally, one should be able to assess this issue from the literature, but we understand that specific information may be lacking here. Here, we employ an arbitrary cut off size of 10 cm in order to evaluate for potential mobility. We appreciate colleagues may want to modify this.

1. Sample is a small fragment which may be stratigraphically mobile, e.g. loose fleck of charcoal or individual bone fragment, with no refitting or spatial indication of its stratigraphic integrity.
2. Sample is <10 cm in maximum dimension with no clear indication of its stratigraphic integrity.
3. Sample is <10 cm in maximum dimension with a high probability of stratigraphic integrity.
4. Sample is >10 cm in maximum dimensions and clearly stratified within an identifiable feature.
5. Sample is >10 cm in maximum dimensions and meaningfully associated with comparable items, e.g. articulated skeleton, discrete organic spread.

If each sample is evaluated by each of these criteria and scored accordingly, a total score of 0–20 on chronometry and 0–16 on interpretation can be obtained, combining for a total 'evaluation score' of 0–36. We suggest that samples scoring 27 or above can be considered reliable enough to use in modelling without further question. On the other hand, those with scores of 9 or less should be rejected as highly unreliable. Those with scores from 10–26 should be accepted with a degree of caution, and ideally modelling should occur both including and excluding dates that fall into this range.

4. Using the evaluation criteria: two examples

Here, we have selected from our preliminary database two French sites for evaluation, one approaching four half-lives and a second at around two half-lives.

4.1. Abri Pataud, Dordogne

Table 1 presents the relevant radiocarbon data for this site. Seven radiocarbon measurements exist for the Gravettian (Perigordian) of the Abri Pataud, which relate to five individual samples. Of these, three AMS measurements were taken at Oxford on the same sample of bone from level 3 lens 2a (OxAs-163, -164, -165). That these are statistically the same age is not surprising. In addition, two other distinct bone samples were measured conventionally from the same lens at Groningen (GrN-4506 and -4721), pertaining to the Perigordian VI. Three dates at least therefore exist for this horizon. We score these measurements at 22 points each,[2] indicating that

[2] Contamination 3: chemical fraction 2: accuracy 2: relevance to human activity 2: relevance to specific archaeological entity 3: quantity/nature of dates 3: materials/stratigraphic issues 2 (erring on the side of caution given lack of data): materials/measurement 3 (assumed, given lack of data): methods and reporting 2.

Table 1
Radiocarbon measurements for the Gravettian (Perigordian VI) of Abri Pataud. Data from Gowlett and Hedges [11], and Vogel and Waterbolk [33,34]

Level	Industry	Lab. number	Result	Method/sample
3, lens 2a	Gravettian [Perigordian VI]	OxA-163	23,180 ± 670	AMS: bone
3, lens 2a	Gravettian [Perigordian VI]	OxA-164	24,250 ± 750	AMS: bone (replication of OxA-163)
3, lens 2a	Gravettian [Perigordian VI]	OxA-165	24,440 ± 740	AMS: bone (replication of OxA-163)
3, lens 2a	Gravettian [Perigordian VI]	GrN-4506	22,780 ± 140	Bone
3, lens 2a	Gravettian [Perigordian VI]	GrN-4721	23,010 ± 170	Bone
3 [lens 2a?]	Gravettian [Perigordian VI]	GrN-1892	21,540 ± 160	Charred bone/ashes ("remaining" fraction of GrN-1864)
3 [lens 2a?]	Gravettian [Perigordian VI]	GrN-1864	18,470 ± 280	Charred bone/ashes ("bone" fraction)

Table 2
Radiocarbon measurements for the Magdalenian of Grotte des Romains. Data from Delibrias et al. [7] and Bridault et al. [5]

Level	Industry	Lab. Number	Result	Method/sample
III	Magdalenian [VI]	Ly-16	14,380 ± 380	AMS: charcoal
III	Magdalenian [VI]	GrA-9709(Lyon-642)	12,690 ± 60	Reindeer bone
Iib	Magdalenian	Ly-356	12,980 ± 240	Bone[s]
Iib	Magdalenian	MC-1215	12,540 ± 400	Shells
Iib	Magdalenian	GrA-9710(Lyon-643)	13,380 ± 60	Reindeer bone
Iib	Magdalenian	GrA-9710(Lyon-432)	12,830 ± 60	AMS: reindeer bone
Iib	Magdalenian	Ly-1307	10,280 ± 630	AMS: charcoal

some caution should be employed in their interpretation and modelling should occur with and without them. Two further conventionally dated samples, sub-samples of the same bulked sample, exist for the same Gravettian/Perigordian VI layer. The divergence between the "bone" fraction (GrN-1864) and the "remaining" fraction (GrN-1892), however, is clearly evident. This may either represent diagenetic contamination of the "remaining" fraction or depleted levels of ^{14}C owing to burning in the "bone" fraction. We are aware that burnt bone is one of the most problematic dating materials, and therefore are duly cautious of this result.

4.2. Grotte des Romains (Ain)

Table 2 presents the relevant radiocarbon data for this site. Two measurements, one AMS and one conventional, exist for level III, and five exist for level IIb of which two are AMS. We scored these at 16 and 14,[3] GrA-9709 scoring less as it is a bulked sample. This indicates that the dates for the Magdalenian VI at this site should be treated with some caution, and that modelling should occur both using and rejecting these results. For the Magdalenian of level IIb the situation is a little different. The five existing dates for this level were measured on three distinct materials including a bone and charcoal pair, which are in general agreement with two (younger) outliers that suggest more than one episode of occupation of this site. We scored these results at 22, 18, 22, 24 and 24 respectively,[4] indicating that, once again, a degree of caution should be employed.

5. Wider use of radiocarbon dates

Following the initial realisation in the 1980s [3,26] that radiocarbon datasets could be used in large-scale demographic modelling in archaeology, such practice is becoming increasingly common (e.g. [4,9, pp. 288–289;15,21]). As noted above, Waterbolk's eighth area of concern was the interpretation of large datasets, and in no other use of chronometric data do the concerns expressed by Waterbolk, elaborated above, become so pertinent (e.g. [22]). The above criteria have been designed as part of a working classification system to audit the absolute age determinations being collected by us for the project mentioned at the start of this paper. By publishing these criteria near the start of this project's life, we hope to incorporate feedback obtained from the

[3] Contamination 3/3: chemical fraction 1/1 (erring on the side of caution): accuracy 3/3: relevance to human activity 2/2: relevance to specific archaeological entity 1/1: quantity/nature of dates 0/0: materials/stratigraphic issues 1/1 (erring on the side of caution given lack of data): materials/measurement 3/3 (assumed, given lack of data): methods and reporting 2/0.

[4] Contamination 3/2/3/3/3: chemical fraction straight 3s: accuracy straight 4s: relevance to human activity straight 2s: relevance to specific archaeological entity straight 2s: quantity/nature of dates straight 3s (erring on the side of caution given 2 outliers): materials/stratigraphic issues straight 2s: materials/measurement 3/0/3/3/3 (assumed, given lack of data): methods and reporting 0/0/0/2/2.

wider archaeological research community into our final auditing of our date database, and work towards a more critically robust use of radiocarbon datasets.

Our own database will contain absolute dates from Europe, the Near East and North Africa, covering the period between 20 and 8 ka (cal) BP. The primary aim of our assessment criteria is to aid our own analysis of human late glacial/early Holocene spatio-temporal patterning seen in this large region, by generating a corpus of consistently and non-intuitively selected absolute dates. However, we have tried to keep our auditing criteria as catholic as possible, so that they may be used and adapted by others. In this paper we throw down the gauntlet.

Acknowledgements

The S2AGES database is supported by the Leverhulme Trust (grant #F00180/1). We are very grateful to Caitlin Buck and Robert Hedges for their valuable comments on a previous draft.

References

[1] M.J. Aitken, Science-Based Dating in Archaeology, Longman, London, 1990.

[2] S.H.R. Aldhouse-Green, P.B. Pettitt, Paviland Cave: contextualizing the Red Lady, Antiquity 72 (278) (1998) 756–772.

[3] A.J. Ammerman, L.L. Cavalli-Sforza, The Neolithic Transition and the Genetics of Populations in Europe, University Press, Princeton, 1984.

[4] J.-P. Bocquet-Appel, P.Y. Demars, Neanderthal contraction and modern human colonisation in Europe, Antiquity 74 (2000) 544–552.

[5] A. Bridault, L. Chaix, G. Pion, C. Oberlin, S. Thiébault, J. Argant, Position chronologique du Renne (Rangifer tarandus L.) à la fin du Tardiglaciaire dans les Alpes du Nord françaises et le Jura mériodional, in: G. Pion (Ed.), Le Paléolithique Supérieur récent: nouvelles données sur le peuplement et l'environnement (Actes de la Table Ronde de Chambéry, Mars 1999), Société Préhistorique Française, Mémoire 28, Paris, 2000, pp. 47–57.

[6] R. Burleigh, K. Mathews, M. Leese, Consensus δ^{13}C values, Radiocarbon 26 (1) (1984) 46–53.

[7] G. Delibrias, M.-T. Guillier, J. Evin, J. Thommeret, Y. Thommeret, Datations absolues des dépôts quaternaires et des sites préhistoriques par la méthode du Carbone 14, in: H. de Lumley (Ed.), La Préhistoire Française, tome 1, CNRS, Paris, 1976, pp. 1499–1514.

[8] F. Djindjian, Les origines du Peuplement aurignacien en Europe, in: L. Bánesz, J.K. Kozlowski (Eds.), Aurignacien en Europe et au Proche Orient (Actes du XII^e Congrès [UISPP], 1991, Bratislava), Institut Archéologique de l'Académie Slovaque des Sciences, Bratislava, 1993, pp. 136–154.

[9] C.S. Gamble, The Palaeolithic Societies of Europe, University Press, Cambridge, 1999.

[10] R. Gillespie, Radiocarbon User's Handbook, Oxford University Committee for Archaeology Monograph 3, 1984.

[11] J.A.J. Gowlett, R.E.M. Hedges, Lessons of context and contamination in dating the Upper Palaeolithic, in: J.A.J. Gowlett, R.E.M. Hedges (Eds), Archaeological Results from Accelerator Dating. Oxford University Committee for Archaeology Monograph 11, Oxford, 1986, pp. 63–71.

[12] R.E.M. Hedges, Radiocarbon dating—50 years on, Education in Chemistry 1997 (1997) 157–164.

[13] R.E.M. Hedges, Radiocarbon dating, in: E. Ciliberto, G. Spoto (Eds.), Modern Analytical Methods in Art and Archaeology, Wiley, London, 2000, pp. 465–502.

[14] R.E.M. Hedges, P.B. Pettitt, On the validity of radiocarbon dates >30 ka BP. Revue d'Archéometrie supplement (proceedings of the 3rd international radiocarbon in archaeology conference, Lyon), 2000, pp. 137–141.

[15] R.A. Housley, C.S. Gamble, M. Street, P.B. Pettitt, Radiocarbon evidence for the Lateglacial recolonisation of Northern Europe, Proceedings of the Prehistoric Society 63 (1997) 25–54.

[16] J.-J. Hublin, F. Spoor, M. Braun, F. Zonneveld, S. Condemi, A late Neanderthal associated with Upper Palaeolithic artefacts, Nature 381 (1996) 224–226.

[17] H. Kitagawa, H. van der Plicht, Atmospheric radiocarbon calibration to 45,000 yr BP: Late Glacial fluctuations and cosmogenic isotope production, Science 279 (1998) 1187–1190.

[18] P.A. Mellars, Archaeology and the population dispersal hypothesis of modern human origins in Europe, in: M.J. Aitken, C.B. Stringer, P.A. Mellars (Eds), The Origin of Modern Humans and the Impact of Chronometric Dating. Philosophical Transactions of the Royal Society, Series B 337 (1280), 1992, pp. 225–234.

[19] P.A. Mellars, The Neanderthal problem continued, Current Anthropology 40 (3) (1999) 341–364.

[20] P.B. Pettitt, Disappearing from the world: an archaeological perspective on Neanderthal extinction, Oxford Journal of Archaeology 18 (3) (1999) 217–240.

[21] P.B. Pettitt, Dating the Mid Upper Palaeolithic: the radiocarbon evidence, in: W. Roebroeks, M. Mussi, J. Svoboda (Eds), Hunters of the Golden Age: The Mid Upper Palaeolithic of Europe. European Science Foundation Workshop on the Palaeolithic Settlement of Europe, Proceedings. Analecta Praehistoria Leidensia, 2000, pp. 21–30.

[22] P.B. Pettitt, A.W.G. Pike, Blind in a cloud of data: problems with the chronology of Neanderthal extinction and early modern human expansion, Antiquity 75 (2001) 415–420.

[23] S. Piggott, The radiocarbon date from Durrington Walls, Antiquity 33 (1959) 289.

[24] C. Renfrew, Before Civilisation, Penguin, London, 1973.

[25] D.A. Richards, J.W. Beck, Dramatic shifts in atmospheric radiocarbon during the Last Glacial period, Antiquity 75 (2001) 482–485.

[26] J.W. Rick, Dates as data: an examination of the Peruvian preceramic radiocarbon record, American Antiquity 52 (1) (1987) 55–73.

[27] F.H. Smith, E. Trinkaus, P.B. Pettitt, I Karavanic, M. Paunovic, Direct radiocarbon dates for Vindija G1 and Velika Pecina Late Pleistocene hominid remains, Proceedings of the National Academy of Sciences (USA) 96 (1999) 12281–12286.

[28] M. Spriggs, The dating of the island Southeast Asian Neolithic: an attempt at chronometric hygiene and linguistic correlation, Antiquity 63 (1989) 587–613.

[29] M. Spriggs, A. Anderson, Late colonisation of East Polynesia, Antiquity 67 (1993) 200–217.

[30] M. Street, M. Baales, B. Weninger, Absolute chronologie des Späten Paläolithikums und des Frühmesolithikums im Nördlichen Rheinland, Archäoligisches Korrespondenzblatt 24 (1994) 1–28.

[31] M. Stuiver, P.J. Reimer, E. Bard, J.W. Beck, G.S. Burr, K.A. Hughen, B. Kromer, G. McCormac, J. Van Der Plicht, M. Spurk, INTCAL98 Radiocarbon age calibration, 24,000–0 cal BP, Radiocarbon 40 (1998) 1041–1083.

[32] T. Van Andel, Middle and Upper Palaeolithic environments and the calibration of ^{14}C dates beyond 10,000 BP, Antiquity 72 (1998) 26–33.

[33] J.C. Vogel, H.T. Waterbolk, Groningen Radiocarbon Dates IV, Radiocarbon 5 (1963) 163–202.

[34] J.C. Vogel, H.T. Waterbolk, Groningen Radiocarbon Dates VII, Radiocarbon 9 (1967) 107–155.

[35] H.T. Waterbolk, Working with radiocarbon dates, Proceedings of the Prehistoric Society 37 (1971) 15–33.

[36] J. Zilhão, F. d'Errico, The chronology and taphonomy of the earliest Aurignacian and its implications for the understanding of Neandertal extinction, Journal of World Prehistory 13 (1999) 1–68.